Smart Systems in Automation Lab ECE-2020

Student Final Project Report: Violence Detection

Name: Pham Lac Duy, Pham Nguyen Hoang Long, Vien Minh Quang Student ID: Your Student ID
Email: Your email EEIT Class: Your EEIT Class

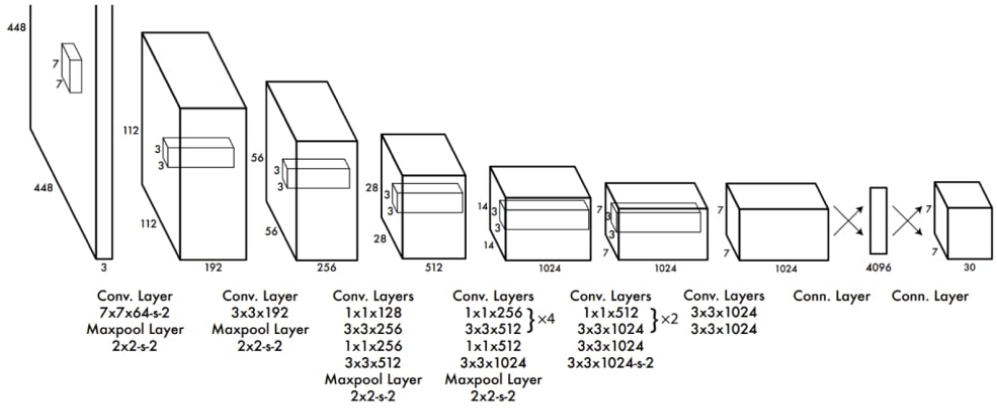
1. Introduction

In this project, we will perform violence detection in real time with the support of YOLOv7.

2. Methodology

- **About the YOLOv7 algorithm:**

The YOLO algorithm takes an image as input and then uses a simple deep convolutional neural network to detect objects in the image. The architecture of the CNN model that forms the backbone of YOLO is shown below.



The Architecture. Our detection network has 24 convolutional layers followed by 2 fully connected layers. Alternating 1×1 convolutional layers reduce the features space from preceding layers. We pretrain the convolutional layers on the ImageNet classification task at half the resolution (224×224 input image) and then double the resolution for detection.

The first 20 convolution layers of the model are pre-trained using ImageNet by plugging in a temporary average pooling and fully connected layer. Then, this pre-trained model is converted to perform detection since previous research showcased that adding convolution and connected layers to a pre-trained network improves performance. YOLO's final fully connected layer predicts both class probabilities and bounding box coordinates.

Link: https://www.v7labs.com/blog/yolo-object-detection#h3

- **Other tools:**

➔ Google Collab, a Web IDE, since it allows running the training process using cloud CPU and GPU.

➔ Label tool: makesense.ai .

- **Collecting data:**

To obtain the data, we recorded a group of volunteer VGU students and ourselves simulating real-life violent and nonviolent situations.

All recorded videos are saved in MP4 video format and then 2 frames per second are extracted using OpenCV in python. We also collected data from the internet.

Our dataset includes 3483 images. There are 2265 images that represent non-violent behaviors and 2779 images that represent violent behavior.

Every violent and non-violent action is manually labeled. As the pictures below

Example of class “Violent”

Example of class “Non-violent”

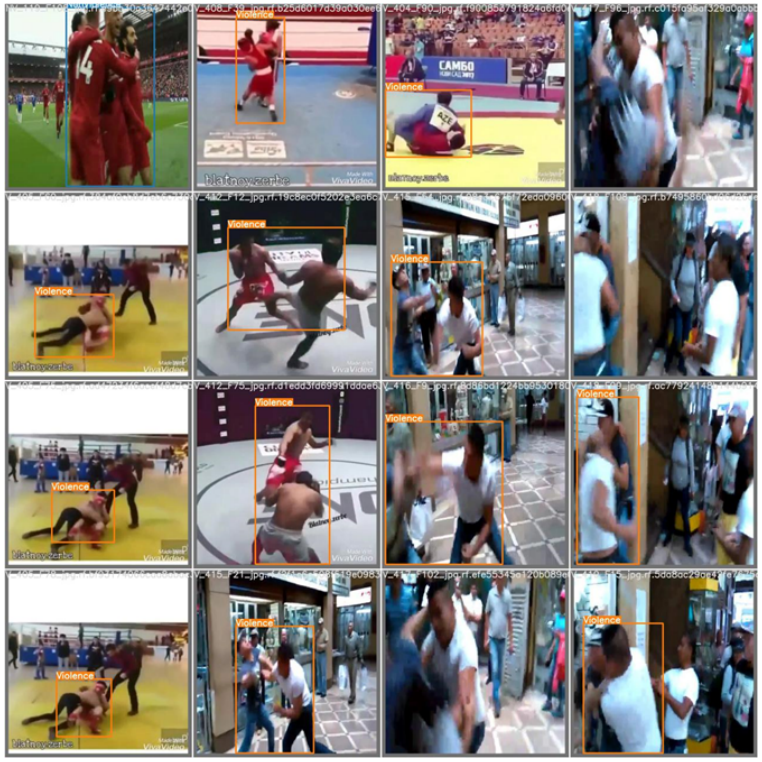
- **System construction:**

Mounting Google Collab to Google Drive in order to manage YOLOv7 training sample and dataset.

Download YOLOv7 repository, install required packages and redirect the path to YOLOv7. In this case, the model is yolov7.yaml .

- **Test image:**

Evaluating test images and displaying inference on all test images



- **Inference on Single Image:**

Then begin to infer and apply Non-Maximum Suppression(NMS). It uses to whittle down many detected bounding boxes to only a few. Once the detector outputs a large number of bounding boxes, it is necessary to filter out the best ones.

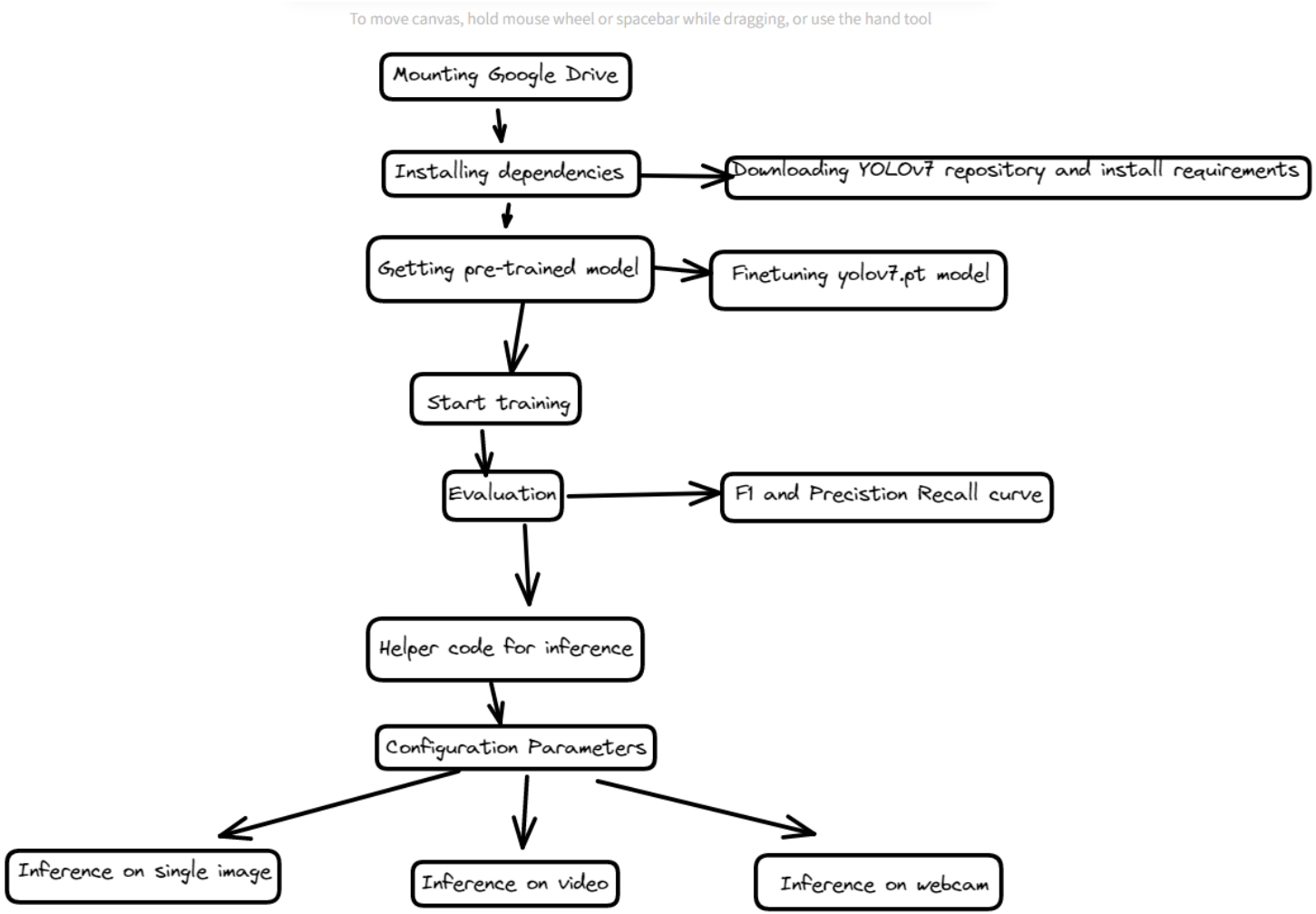
- **Inference on Video:**

We can infer video from the local device or download video from Google Drive.

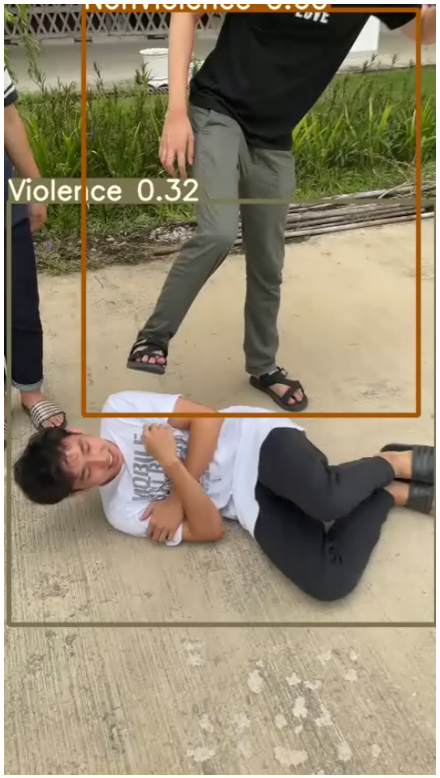
➔ First, we initialize the video object and set the Video information

➔ We create an object for writing video output. Then after initializing the model and loading FP32 model, we set the model for inference.

➔ It also needs Non-Max suppression (NMS). The system will detect per class and decide the object in that frame is violent or non-violent. Then release the output.


To move canvas, hold mouse wheel or spacebar while dragging, or use the hand tool
Mounting Google Drive
Installing dependencies
Downloading YOLOv7 repository and install requirements
Getting pre-trained model
Finetuning yolov7.pt model
Start training
Evaluation
F1 and Precistion Recall curve
Helper code for inference
Configuration Parameters
Inference on single image
Inference on video
Inference on webcam

3. Result:



- **Pros and cons:**

Pros

- ➔ Can process images and videos
- ➔ Real-time detection with webcam
- ➔ User-friendly interface
- ➔ Can be used on local network

Cons

- ➔ Accuracy is reduced when applied to photos or videos with an excessive number of people
- ➔ The types of violent behavior are not yet diversified, especially armed violence
- ➔ Cannot go online

4. Future work:

More diverse image sources

Armed violence

Online website

Communication with local authorities and the owners of the security system via SMS or email.

- **Conclusion:**

By taking violent object detection from photographs into consideration, the cutting-edge YOLO (you only look once) deep-learning-based algorithm's detection capabilities have been examined. The YOLOv7 model was trained with violent and nonviolent photos from our own dataset and real-world events for the model search component to

improve the performance of the YOLOv7 model and enable it to detect violence with AP values as high as 86.6 percent.